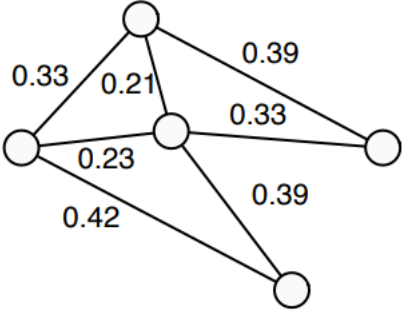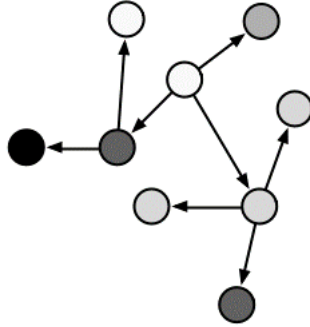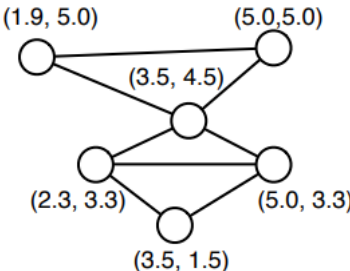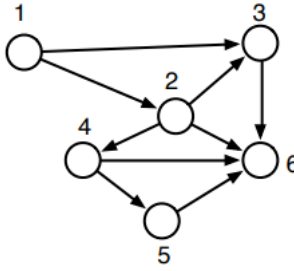# GPBR Exercise 1

Name: Soi Zhi Wen
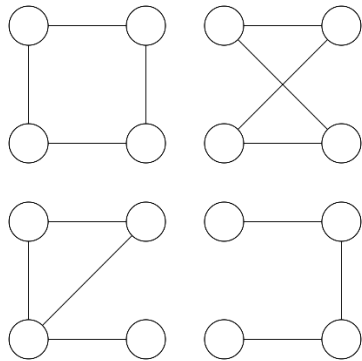
Matriculation Number: 22-132-245

1.

Advantages	Disadvantages
Graphs have the ability to capture complex relationships and dependencies between data points that could be challenging to represent using vectorial representations.	Scaling graph-based pattern representation to very large datasets can be challenging, whereas vectorial representations can be easily scaled to handle large datasets.
Graph-based pattern representation is typically highly interpretable because the graph explicitly represents the relationships among data points, while vectorial representations may lack interpretability since the relationships among data points are not explicitly represented.	Constructing and analyzing graph-based pattern representation can be computationally demanding, particularly for large and dense graphs. On the other hand, vectorial representations can be highly efficient to compute, making them well-suited for large datasets.

2.

Weighted graph

Directed with labeled nodes

Graph with unique node labels

Ordered graph

3. (a) $N(4,3) = 2$

   (b)

   Total number of non-isomorphic graphs of size 4 is 4.

4.

   $$
   \begin{array}{ccc}
   1 & 0 & 1 \\
   0 & 1 & 0 \\
   1 & 0 & 1
   \end{array}
   $$

   $f(v_1) = w_1$
   $f(v_2) = w_2$
   $f(v_3) = N/A$
   $f(v_4) = w_4$
   $f(v_5) = w_3$
   $f(v_6) = N/A$

   $f(v_1) = x_1$
   $f(v_2) = x_4$
   $f(v_3) = x_2$
   $f(v_4) = x_5$
   $f(v_5) = x_6$
   $f(v_6) = x_3$

   $f(w_1) = x_1$
   $f(w_2) = x_4$
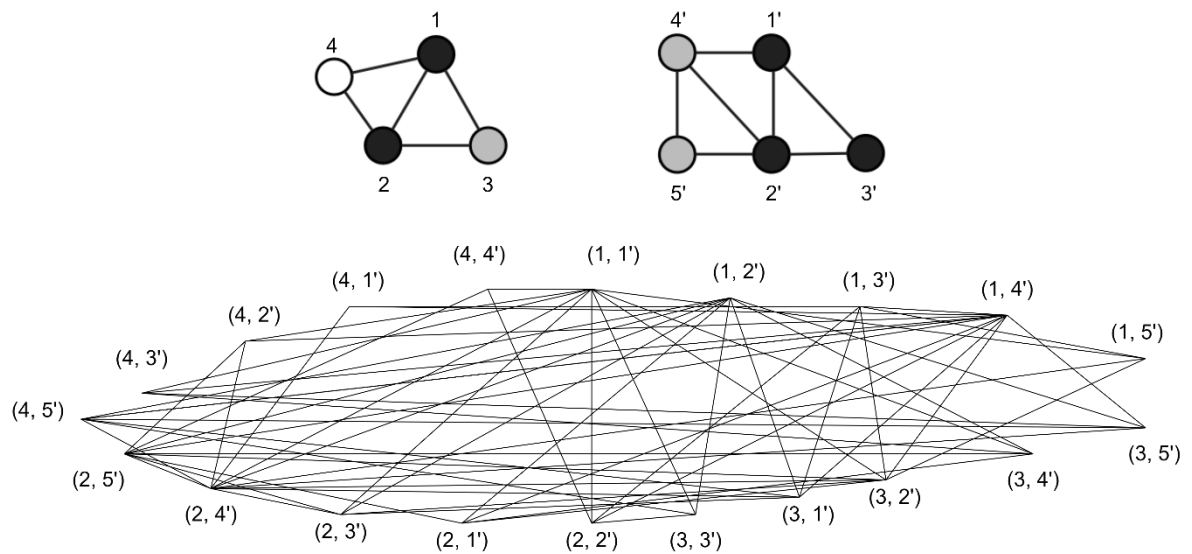 $f(w_3) = x_6$
   $f(w_4) = x_5$
   $f(w_5) = N/A$
   $f(w_6) = N/A$

5.

The maximum clique is 3.

6. (a) Shortest-path kernels on graphs
In the paper, a technique for identifying frequent subgraphs in a database of graphs is presented, where each subgraph is represented as a labeled graph denoted as $g = (V, E, \mu, \nu)$. The nodes of the graph correspond to the various types of subgraphs that are being sought, with each node being labeled with attributes specific to the type of subgraph it represents. Similarly, the edges of the graph represent relationships between the different types of subgraphs, and each edge is labeled with attributes describing the relationship between the nodes it connects. The labeling functions μ and ν are utilized to capture the semantics of subgraphs and their relationships, enabling the representation of subgraph patterns and relationships in a flexible and comprehensive manner that can support efficient subgraph mining algorithms.

(b) Graph Similarity Features for HMM-Based Handwriting Recognition in Historical Documents
The article discusses a technique for recognizing handwriting in historical documents by utilizing graph similarity features. The writing patterns are represented as a graph $g = (V, E, \mu, \nu)$, where each node represents a stroke and is labeled with attributes such as position, direction, and curvature of the stroke. Edges represent the temporal order of the strokes, and each edge is labeled with attributes such as the distance and angle between the corresponding strokes. The graph representation allows for a flexible and detailed representation of the handwriting patterns, which can be used for accurate handwriting recognition. Graph similarity features are computed by comparing the corresponding graphs of the training and test data.

(c) Malware Classification based on Call Graph Clustering
The paper proposes a graph theory-based method for analyzing complex networks. The complex network is represented as a graph $g = (V, E, \mu, \nu)$, where nodes

represent the entities or components of the network, such as proteins or websites, and edges represent the connections or interactions between the nodes. Each node and edge is labeled with a set of attributes that capture its properties, such as degree, centrality, weight, or strength. The node and edge labeling functions μ and ν map each node and edge to their corresponding set of attributes. The graph representation enables the analysis of the structural properties of complex networks in various fields, including biology, sociology, and computer science, to understand their behavior and predict their evolution.

(d) A Graph Matching Based Approach to Fingerprint Classification Using Directional Variance

The paper presents a molecular graph approach for graphically representing chemical compounds, enabling a flexible and powerful analysis of their structure and properties. The approach employs a graph $g = (V, E, \mu, \nu)$ where the nodes represent the atoms in the compound and the edges represent the bonds between the atoms. Each node is labeled with a symbol corresponding to the atomic element of the atom, such as "C" for carbon and "O" for oxygen. Additionally, each node is labeled with a set of attributes capturing properties such as position and charge. Each edge is labeled with a symbol representing the type of bond between the corresponding atoms, such as single, double, or triple bonds. The node labeling function μ maps each node to its corresponding symbol and set of attributes, while the edge labeling function ν maps each edge to its corresponding bond type symbol. This graph representation provides a flexible and powerful tool for analyzing the structure and properties of chemical compounds, making it possible to predict their interactions and behavior. It has applications in various fields such as chemistry, materials science, and drug discovery.